文章编号:1673-2677(2026)01-0114-11 DOI:10.12388/j.issn.1673-2677.2026.01.013
引用:苏斌,李俊超,朱晨,等.机器学习驱动的二氧化碳气水交替驱开发策略优化方法[J].新疆石油天然气,2026,22(1):114-124.
Cite:SU Bin,LI Junchao,ZHU Chen,et al. An optimization method for CO_2 water-alternating-gas (WAG) development strategy driven by machine learning[J]. Xinjiang Oil & Gas,2026,22(1):114-124.

机器学习驱动的二氧化碳气水交替驱开发策略优化方法

苏斌,李俊超,朱晨,李纪新
(西安石油大学,陕西西安 710065)

摘 要:数值模拟方法在CO_2气水交替驱(CO_2-WAG)开发策略优化中虽然具有较高的预测精度,但在多参数组合优化和全局优化中因计算量巨大使其效率受限,难以满足复杂油藏开发方案的高效优化需求。为了解决这一问题,提出了一种基于随机森林模型的数值模拟替代方法,并结合全局优化框架显著提升了开发策略优化的计算效率与预测精度。通过改进的K-Fold交叉验证方法优化模型超参数,构建了能精准捕捉输入参数与开发效果之间复杂非线性关系的替代模型。以随机森林模型的测试集拟合优度(R^2)和均方根误差(RMSE)为主要评价指标,验证了模型的预测可靠性,并通过与数值模拟结果对比进一步论证了方法的可行性和精确性。优化中结合循环遍历法探索增油量与CO_2埋存率之间的权衡,提出了多目标下的最优注采策略。优化结果表明,与传统开发方案相比,基于随机森林模型优化后的开发方案能够在显著提高累计产油量的同时增加CO_2埋存量并延迟气窜时间。在某致密油藏的实际模型中,相比样本最优方案,优化方案使累计产油量提升8.6%,CO_2埋存量增加14.5%,气窜时间延迟6年,显著改善了驱替效果和提高了开发效益。

关键词:CO_2-WAG;致密油藏;随机森林;主控因素;机器学习
中图分类号:TE357.45;TE319 **文献标识码**:A

An Optimization Method for CO_2 Water-Alternating-Gas (WAG) Development Strategy Driven by Machine Learning

SU Bin,LI Junchao,ZHU Chen,LI Jixin
(Xi'an Shiyou University,Xi'an 710065,Shaanxi,China)

Abstract:Although numerical simulation exhibits high predictive accuracy in optimizing development strategies for CO_2 water-alternating-gas (CO_2-WAG) flooding,its limited efficiency cannot facilitate high-efficiency optimization of development plans for complex reservoirs due to the enormous computational burden involved in multi-parameter combinational and global optimization. To address this issue,a surrogate modeling approach based on the random forest algorithm is proposed and integrated with a global optimization framework to greatly enhance computational efficiency and predictive accuracy. By employing an improved K-Fold cross-validation method for hyperparameter optimization,a surrogate model capable of accurately capturing the complex nonlinear relationships between input parameters and development outcomes is constructed. The reliability of the model is validated using the test set of goodness of fit(R^2) and root mean square error (RMSE),and its feasibility and precision are further demonstrated through comparisons with numerical simulation results. During the optimization process,an iterative traversal method is employed to explore the trade-off

基金项目:国家自然科学基金青年基金项目"陆相中低熟页岩油原位转化开发裂缝动态演化机理研究"(52204047);国家自然科学基金项目"井下多化学反应生烃放热协同汽大幅度提高稠油采收率基础问题研究"(U23B6003)。
第一作者:苏斌(1997-),西安石油大学机械工程学院在读硕士,目前从事油气藏数值模拟和地应力模拟研究。(E-mail)subinsjl@163.com
通讯作者:李俊超(1987-),2015年毕业于北京大学力学专业,博士,教授,目前从事油气藏数值模拟及相关软件研发和教学工作。(E-mail)lijunchao@xsyu.edu.cn

between incremental oil recovery and CO_2 storage efficiency, leading to the proposal of an optimal injection-production strategy meeting multiple objectives. The optimization results reveal that compared with conventional development schemes, the random forest algorithm optimized strategy significantly increases cumulative oil production, enhances CO_2 storage efficiency, and delays gas breakthrough. For a real tight reservoir model, the optimized scheme delivers an increase of 8.6% in cumulative oil production, an improvement of 14.5% in CO_2 storage, and a six-year delay in gas breakthrough, demonstrating significant enhancements in displacement performance and development benefits.

Key words: CO_2-WAG; tight reservoir; random forest; main controlling factor; machine learning

CO_2驱油技术是一种针对非常规油藏的有效增产手段[1]。该技术不仅能显著提高原油采收率，还兼具捕集和封存温室气体的功能，是碳捕获、利用与封存-提高采收率（CCUS-EOR）的重要发展方向[2-3]。在评估CO_2驱油效果时，通常以原油增产量和CO_2埋存率作为关键评价指标[4]。但在实际应用中，由于CO_2黏度较低、流动性强，且油藏普遍存在非均质性，CO_2驱油过程中极易发生黏性指进和气窜现象，造成流度比失衡，严重影响开发效果[5]。气窜的发生可能导致注入的CO_2绕过目标油层，快速穿透至生产井，不仅降低波及体积和驱替效率，还会引发生产井气油比急剧上升、产液量下降等问题，进而增加了开发难度和潜在安全风险[6-8]。CO_2气水交替驱（CO_2-WAG）技术被认为是延缓CO_2驱油过程中气窜问题的有效手段[9]。该技术通过在同一注采系统中交替注入CO_2与水，利用气-水两相流体协同作用形成复合驱替机制，从而改善驱替效率、扩大波及体积并有效延缓气体突破时间。与常规气驱相比，CO_2-WAG技术不仅能够显著提高微观驱替效率，还能通过控制流体流动稳定性提升宏观驱替效率[10]。然而，非常规油藏普遍存在低孔隙度、低渗透率和复杂孔喉结构等地质特征，导致注入的CO_2和水在地层中的运移规律复杂，驱替表现受多种因素联合作用的影响，进而加剧了驱油效果的不确定性[11]。目前，CO_2-WAG驱替过程中各个开发参数之间的耦合影响机制尚未完全明确，各主控因素的识别与优化仍面临较大挑战。因此，在实施CO_2-WAG过程中优化开发参数和准确预测生产性能仍然是研究的重点。

传统的油藏预测方法，如经验公式法、油藏工程分析法及数值模拟法，已在实际应用中得到了广泛推广[12]。尽管经验公式法和油藏工程法在特定地质条件和操作环境下具有一定的有效性，但其适用范围通常较为有限，难以适应复杂多变的工程实践需求[13-14]。数值模拟法能详尽揭示多相流体在多孔介质中的流动特性，是传统油藏开发预测的重要基础[15]。虽然数值模拟法在单一开发方案的效果预测中具有较高精度，但其在处理多参数组合优化及全局优化问题时，往往因计算量庞大而导致效率低下，难以满足复杂油藏开发的实际需求。特别是在CO_2-WAG驱的多目标优化过程中，随着油藏强非均质性特征及多参数条件的引入，数值模拟的计算负担呈指数增长，单个模拟案例的运行时间可能长达数小时甚至数天，这使得其在实时油藏管理及多参数优化中的应用变得难以实现[16,17]。因此，开发一种高效且可靠的替代方法，既能保持数值模拟的预测精度，又能显著提升参数优化效率，已成为当前CO_2-WAG驱油藏开发研究的关键方向和亟需解决的核心问题。

近年来，人工智能与数据科学的快速发展激发了机器学习在石油工程领域中的广泛应用潜力[18]。机器学习模型在处理复杂非线性系统、加速历史拟合进程及克服传统数值方法在收敛性上的局限性方面展现出显著优势[19-22]。借助高质量的数值模拟数据，机器学习模型能够高效学习油藏系统的内在动态特征，从而实现快速的生产预测与开发方案优化[23]。You等[24]提出了一种基于机器学习的优化流程，将人工神经网络（ANN）与粒子群优化算法（PSO）相结合，成功提升了CO_2-WAG项目中的原油采收率及CO_2封存效率。Vo Thanh等[25]对四种机器学习模型：广义回归神经网络（GRNN）、自适应共轭神经网络-最小二乘法（CFNN-LM）、自适应共轭神经网络-贝叶斯回归（CFNN-BR）及极端梯度提升法（XGBoost）进行了评估，用于预测CO_2泡沫驱油过程中的原油采收率因子（ORF），研究结果显示实验成本和时间均显著降低。此外，多项研究将机器学习方法与优化技术相结合，在CO_2-WAG参数优化与生产预测方面取得了显著进展[26-27]。尽管已取得诸多成果，机器学习方法在实际应用中仍面临诸如对初始值敏感、超参数调优复杂、高维数据处理困难及多目标优化能力有限等挑战。

作为一种非参数型集成机器学习方法，随机森林算法具备对高维、非线性及含噪声数据的良好鲁棒性，且预测精度较高，近年来已被广泛应用于油藏工程复杂情况的预测与分析任务[28]。通过集成大量决策树模型，随机森林算法能够有效识别输入变量与输出指标之间的复杂非线性关系，特别适用于CO_2-WAG等复杂驱替过程的建模与分析。但随机森林模型的性能在很大程度上依赖于超参数的合理设定，例如决策树的数量、节点分裂准则、树的最大深度等[29]。K-Fold交叉验证方法是一种常用的模型评估

与超参数调优方法，在数据量有限或数据分布复杂的情况下能够显著提升模型的泛化能力与预测精度。在复杂的油藏数值模拟研究中，传统K-Fold方法可能难以有效应对数据分布不均或特定子集对模型性能的显著影响[30]。针对传统数值模拟方法在优化效率方面的局限性，本文提出了一种基于随机森林模型的数值模拟替代方法，并结合全局优化框架，显著提升了CO_2-WAG驱开发策略优化的效率与精度。通过改进的K-Fold交叉验证方法对随机森林模型的超参数进行优化，构建了一个能够准确捕捉多输入参数与开发效果之间复杂关系的代理模型。以累计产油量和CO_2埋存率为优化目标，提出了多目标条件下的最优注采策略。优化结果表明，所提出的方法不仅显著提升了开发效果，还有效延长了气窜时间，增强了驱替过程的稳定性。

1 数学模型

为准确描述致密油藏中的CO_2-WAG驱开发过程，需结合实际油藏特征与驱替机理对复杂物理现象进行合理简化与假设，以建立适用于致密油藏的CO_2-WAG驱油数学模型。本文对模型构建给出如下基本假设。

1）研究对象为致密砂岩油藏，其地质特性表现为低渗透、孔隙结构复杂、具弱可压缩性。

2）油藏流体视为油、气、水三相多组分流动体系，允许CO_2在气相与油相之间发生溶解，但忽略其他组分在相界面上的质量交换。

3）流体在储层中的渗流遵循达西定律，考虑重力和毛管压力的影响，忽略温度变化、化学反应及矿物反应对渗流的影响。

4）注入气体为CO_2，生产井采用定井底流压生产，假设井筒内流体瞬时达到稳定流动状态。

基于上述假设，本文采用多相多组分渗流理论建立CO_2-WAG驱油数学模型。对任意组分j在多孔介质中的质量守恒满足以下控制方程：

$$\frac{\partial}{\partial t}\left(\phi\sum_{\alpha}X_{\alpha j}\rho_{\alpha}S_{\alpha}\right)+\nabla\cdot\sum_{\alpha}X_{\alpha j}\rho_{\alpha}u_{\alpha}+Q_{j}=0 \quad (1)$$

式中：ϕ为孔隙度，无量纲；$X_{\alpha j}$为组分j在相α中的质量分数，无量纲；ρ_{α}为相α的密度，kg/m³；S_{α}为相α的饱和度，无量纲，满足油气水三相饱和度之和为1；u_{α}为相α的达西流速，m/s；Q_j为组分j的井源汇项，kg/(m³·s)。

储层中各相的流动遵循达西定律，相α的渗流速度由下式给出：

$$u_{\alpha}=-\frac{kk_{r,\alpha}}{\mu_{\alpha}}(\nabla P_{\alpha}-\rho_{\alpha}g) \quad (2)$$

式中：k为绝对渗透率，mD；$k_{r,\alpha}$为相α的相对渗透率，无量纲；μ_{α}为相α的黏度，Pa·s；P_{α}为相α的压力，Pa；g为重力加速度，m/s²。

对于多相渗流理论，组分j在不同相之间的分配遵循相平衡条件：

$$f_{j,o}=f_{j,g},\quad f_{j,g}=f_{j,w} \quad (3)$$

式中：$f_{j,o}$，$f_{j,g}$，$f_{j,w}$分别为组分j在油相、气相和水相中的逸度，Pa。

任一相的逸度$f_{j,\alpha}$可通过Peng-Robinson状态方程[31]计算：

$$f_{j,\alpha}=\phi_{j,\alpha}x_{j,\alpha}P_{j} \quad (4)$$

式中：$\phi_{j,\alpha}$为逸度系数，依赖于压力、温度和组分浓度，无量纲；$x_{j,\alpha}$为组分j在相α中的摩尔分数，无量纲；P_j为油藏压力，Pa。

在CO_2-WAG驱油数值模拟过程中，流体密度、黏度的变化以及CO_2的溶解行为是影响模拟精度和驱油效果预测的关键因素，可通过以下公式描述。

流体相α的密度ρ_{α}是其内部各组分密度的加权结果，通过组分的摩尔分数和摩尔质量计算：

$$\rho_{\alpha}=\frac{\sum_{j}x_{j,\alpha}M_{j}}{V_{m,\alpha}} \quad (5)$$

式中：M_j为组分j的摩尔质量，kg/mol；$V_{m,\alpha}$为相α的摩尔体积，m³/mol。

CO_2的注入会改变油相和气相的组分比例，从而导致流体黏度的动态变化。CO_2溶解于原油中会显著降低油相黏度，增强原油的流动性；气相中CO_2的黏度相对较低，会导致气相在流动过程中更容易发生窜流。流体相α的黏度μ_{α}通常采用加权平均的混合法则计算[32]：

$$\mu_{\alpha}=\sum_{j}x_{j,\alpha}\mu_{j,\alpha} \quad (6)$$

式中：$\mu_{j,\alpha}$为组分j在相α中的黏度贡献值，Pa·s。

气相和油相间的CO_2溶解行为遵循亨利定律[33]，即CO_2在油相中的溶解度与气相中CO_2的分压成正比：

$$C_{CO_2}^{o}=H_{CO_2}p_{CO_2}^{g} \quad (7)$$

式中：$C_{CO_2}^{o}$为油相中溶解的CO_2浓度，mol/m³；H_{CO_2}为CO_2在原油中的亨利系数，mol/(m³·Pa)；$p_{CO_2}^{g}$为气相中CO_2的分压，Pa。

此外，毛管压力作用下各相压力之间的关系描述如下：

$$P_c^{ow} = P_o - P_w, P_c^{go} = P_g - P_o \tag{8}$$

式中：P_c^{ow}为油水毛管压力，Pa；P_c^{go}为气油毛管压力，Pa；P_o、P_w、P_g分别为油相、水相、气相压力，Pa。

2　随机森林模型的建立

随机森林（RF）是一种经典的集成机器学习算法。其核心思想是通过构建多个相互独立的决策树分类器或回归器，以投票或加权平均等集成方式获得整体模型的预测结果，从而有效降低单个树模型的预测偏差与方差，显著提升预测精度和泛化能力[34]。在实际应用中，随机森林模型的性能高度依赖于超参数的合理选择，主要包括决策树数量、最大树深度、节点最小分裂样本数等。为提升模型性能，通常需采用系统化、自动化的超参数优化方法，例如交叉验证方法，以实现模型超参数的精准调优。

2.1　数据样本获取与预处理

高质量的数据样本是建立可靠随机森林模型的基础。本文通过分析致密油藏CO_2-WAG驱油过程，明确了影响开发效果的关键因素，并剔除冗余或低相关性变量，最终筛选出生产压差、单轮次注入时间、注水速度、注气速度及注水时间百分比五个核心参数，作为随机森林模型的输入变量。图1展示了随机森林模型预测CO_2-WAG驱油开发效果的整体工作流程。模型以这五个核心变量构建数据集，并通过生成多个决策树对其进行分析和预测。在此过程中，每棵决策树独立完成对开发效果的预测，生成多个预测结果。随后，模型利用均方根误差（RMSE）评估预测的准确性，确保结果的可靠性和稳定性。最终，随机森林模型通过综合所有决策树的预测结果，提供对CO_2-WAG驱油开发效果的整体评估。

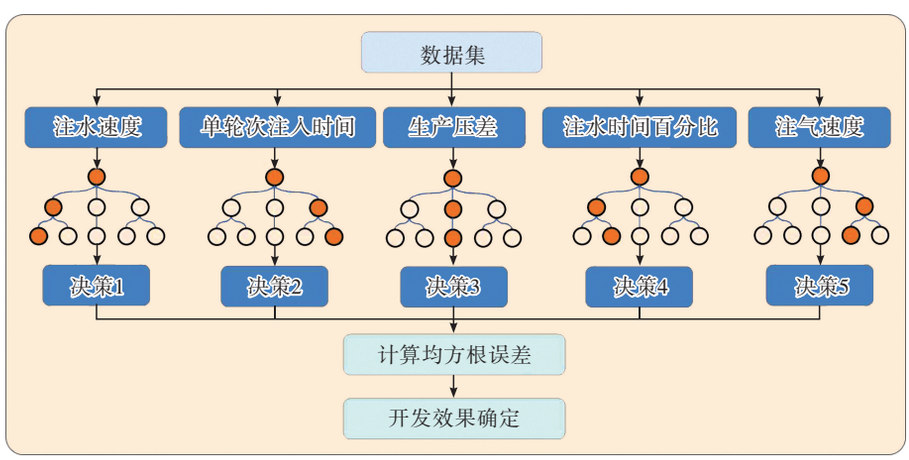


图1　随机森林模型预测CO_2-WAG驱油开发效果的工作流程

Fig. 1　Workflow of random forest model for predicting CO_2-WAG flooding development performance

由于各输入参数的量纲与数值范围存在明显差异，为避免某些变量对模型结果的过度影响，需对原始数据进行归一化处理。本文采用线性极值归一化法将所有数据映射到[0，1]区间：

$$X_i^* = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{9}$$

式中：X_i^*为第i个数据的归一化值；X_i为第i个数据的原始值；X_{max}和X_{min}分别为数据集中的最大值和最小值。

归一化处理完成后，采用拉丁超立方抽样方法（LHS）在参数空间内生成多个组合样本，并利用油藏数值模拟软件分别计算每个样本对应的CO_2埋存率和增油量指标。CO_2埋存率r_s定义为注入CO_2总量与产出CO_2总量之差占注入CO_2总量的比例：

$$r_s = \frac{G_{inj} - G_{prod}}{G_{inj}} \tag{10}$$

式中：r_s为CO_2埋存率，无量纲；G_{inj}为累计注入CO_2量，t；G_{prod}为累计产出CO_2量，t。

原油增产指标r_r定义为CO_2-WAG驱开发方案下相较于常规开发方式的增产效果，采用归一化处理形式表达：

$$r_r = \frac{(N_p - N_{p0}) - N_{p,min}}{N_{p,max} - N_{p,min}} \tag{11}$$

式中：r_r为增油量归一化指标，无量纲；N_p为CO_2-WAG驱条件下累计原油产量，t；N_{p0}为常规开发方式累计原油产量，t；$N_{p,max}$、$N_{p,min}$分别为原油增产量的最大值和最小值，t。

为同时满足CO_2-WAG驱开发过程中对提高产油量和增加CO_2埋存率的综合需求，进一步构建综合评价指标R：

$$R = w_s r_s + w_r r_r \tag{12}$$

式中：R为综合评价指标，无量纲；w_s、w_r分别为CO_2埋

存率与增油量的权重系数，无量纲，满足$w_s + w_r = 1$。

2.2 超参数优化方法

随机森林模型的预测性能在很大程度上取决于其超参数设置的合理性。为避免模型出现过拟合或欠拟合问题，通常采用K-Fold交叉验证方法进行超参数优化。但传统的K-Fold方法在实际应用中存在一定局限性，如参数选择过于依赖经验或主观判断，可能导致模型性能波动较大，优化效果不稳定。针对上述问题，本文提出一种改进的K-Fold交叉验证方法，以降低超参数选择过程中对经验和主观因素的依赖程度。改进方法的核心内容如图2所示，针对数据集划分比例、最小叶子节点数和决策树数量等关键超参数，分别采用算术差分法、网格搜索法和指数递增法进行系统优化。通过对上述超参数进行组合优化，并基于随机森林模型的RMSE评价指标开展性能评估，能够有效提升模型预测精度和稳定性，显著增强模型的拟合能力与泛化表现。

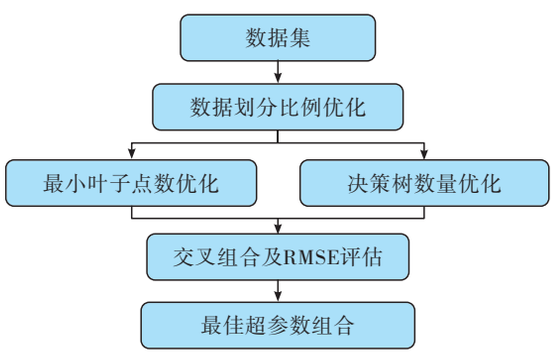


图2 改进的K-Fold交叉验证超参数优化流程示意图

Fig. 2 Flowchart of the improved K-Fold cross-validation method for hyperparameter optimization

2.2.1 数据集划分比例优化

为获得稳健的训练-测试数据划分方案，使用等差序列法对训练集占比进行参数寻优。具体计算方法见公式(13)～公式(16)。

$$v_{tr} = a + (n-1)\Delta \tag{13}$$

$$v_{te} = 1 - v_{tr} \tag{14}$$

式中：v_{tr}为训练集比例；v_{te}为测试集比例；a为训练集比例初始值；Δ为训练集比例步长；n为训练集比例取值序号。

对应具体的训练集和测试集算例数计算方法：

$$N_{tr} = v_{tr}N \tag{15}$$

$$N_{te} = v_{te}N \tag{16}$$

式中：N_{tr}为训练集算例数；N_{te}为测试集算例数；N为数据集总样本数。

2.2.2 最小叶子节点样本数优化

为避免决策树叶子节点过度细分引起的过拟合问题，采用网格搜索法优化最小叶子节点样本数：

$$L_i = L_{i-1} + d, i = 2,3,\cdots, O \tag{17}$$

式中：L_i为第i次取值的最小叶子节点样本数；L_1为初始取值；d为搜索步长；O为取值方案总数。

2.2.3 决策树数量优化

决策树数量过少可能导致预测精度不足，而树的数量过多则会引起计算效率降低和模型冗余。为合理确定决策树数量，采用指数递增法进行优化搜索：

$$T_j = T_{min}m^j, j = 0,1,2,\cdots,M-1 \tag{18}$$

式中：T_j为第j次取值对应的决策树数量；T_{min}为初始决策树数量取值；m为递增倍数(增长底数，$m > 1$)；m^j为m的j次幂，表示在第j次取值时决策树数量相对于T_{min}的增长倍数；M为决策树数量的取值方案总数。

通过上述改进的交叉验证方法，可对随机森林模型的超参数进行系统化、精细优化完善，有效提升模型在致密油藏CO_2-WAG开发预测中的拟合能力与泛化表现，为后续主控因素识别及开发策略优化奠定基础。

3 实例应用

为验证本文提出的随机森林模型在致密油藏CO_2-WAG驱开发策略优化中的有效性，选取我国某致密油藏实际区块作为研究算例进行分析。该数值模型的建立与求解遵循了第1章所阐述的描述CO_2-WAG驱多相多组分渗流规律的数学模型，综合考虑了CO_2在油水两相中的溶解、流体黏度动态变化、毛管压力及重力等关键物理效应。在此理论框架下建立了具体的储层数值模型，如图3所示。所选区块面积约为2.9 km^2，采用角点网格对储层进行三维划分，平面网格尺寸30 m×30 m，纵向根据实际地质分层划分144个小层，平均单层厚度0.85 m，总网格数约23×10^4个。该油藏地质参数较为复杂，初始油藏压力45 MPa，孔隙度3%~10%，渗透率0.03~3.26 mD。区块内共布置7口垂直井，包括4口生产井(WELL-P1至WELL-P4)和3口注入井(WELL-I1至WELL-I3)。开发模式设定为：注入井交替注入CO_2和水，生产井以定井底流压方式开采，模拟总时长为20年。

根据该实例的生产开发特征，CO_2埋存率和增油量综合指标权重系数分别设为0.75和0.25。运用拉丁超立方抽样方法对参数进行随机取值，建立了485个参数组合及对应算例，利用油藏数值模拟软件计算了CO_2埋存率和增油量的综合指标，作为模型样本集。

3.1 模型优化与性能验证

3.1.1 超参数取值优化策略

随机森林模型的性能高度依赖于数据的质量与连续性。为确保输入数据的代表性和连续性，在分析

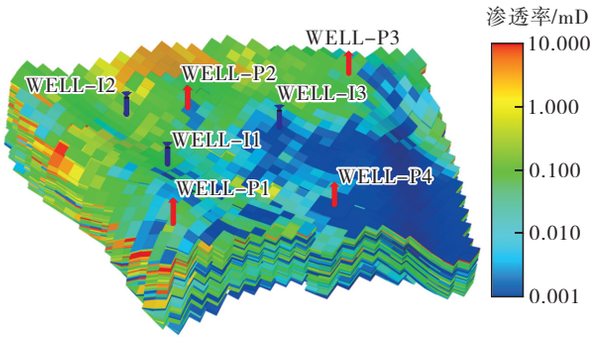


图3　致密油藏储层数值模型示意图

Fig. 3　Numerical reservoir model for a tight oil reservoir

CO_2-WAG驱替过程中主控因素作用机理的基础上，明确了五个关键参数作为输入变量，分别为生产压差、注气速度、注水速度、单轮次注入时间和注水时间百分比。其中，注水时间百分比和单轮次注入时间两个连续参数用于共同表征传统的水气周期比（离散型参数）。各关键参数的具体取值范围见表1。

表1　关键参数取值范围

Tab. 1　Ranges of key parameters

变量	取值范围
生产压差/MPa	10~20
注气速度/(10^3 $m^3\cdot d^{-1}$)	5~20
注水速度/($t\cdot d^{-1}$)	20~40
注水时间百分比/%	0~99
单轮次注入时间/d	90~450

为获得数据样本，首先利用拉丁超立方抽样方法在上述参数空间内随机生成485个参数组合，然后采用油藏数值模拟软件对各组合进行模拟计算，获得每个方案对应的CO_2埋存率与增油量。根据本文实例的生产开发要求，CO_2埋存率和增油量综合指标权重系数分别设为0.75和0.25。随机森林模型的性能受超参数设置影响显著，采用改进的K-Fold交叉验证方法对随机森林模型的决策树数量、最小叶子节点样本数以及训练-测试集划分比例等关键超参数进行优化。具体超参数优化取值方案见表2。通过不同参数组合获得多组模型并计算各自的RMSE，选取最优组合。

表2　超参数取值方案

Tab. 2　Hyperparameter value sets

超参数	取值集合
最小叶子节点数	1,3,5,7,9,11
决策树数目	5,25,125,625
训练集划分比例系数	1,3,5,7,9
测试集划分比例系数	9,7,5,3,1

3.1.2　模型优化结果与性能评估

将上述超参数组合代入随机森林模型，进行交叉评估与优化。不同超参数组合对应RMSE如图4所示。随着最小叶子节点数的逐渐增加，模型出现欠拟合的趋势逐步增强，预测误差相应增大；当决策树数量由5增加到625时，模型预测误差先快速降低，随后趋于稳定，这表明过多的决策树数量对模型性能的提升效果逐渐减弱。表明决策树数量的增加可以有效降低预测误差，但模型数量过多会导致计算成本明显增高。因此，需要在降低误差与计算成本之间取得合理的平衡。此外，对训练集与测试集的划分比例进行优化分析发现，增加训练集的数据比例能显著提升模型对数据的拟合程度和预测稳定性。通过对比分析RMSE与综合考虑计算成本，最终确定最优超参数组合为：最小叶子节点数为1，决策树数量为625，训练集与测试集比例为9:1。

采用上述最优超参数组合，进一步对随机森林模型的拟合性能与泛化能力进行了深入分析与验证。优化后模型拟合效果如图5所示。优化后的模型在训练集数据上的拟合效果表现优异，在训练集上的拟合优度（R^2）达到0.98，测试集上的拟合优度也达到0.93，反映出模型具备较高的拟合精度和稳定的泛化表现。图6为模型均方根误差变化曲线。优化后模型的RMSE趋于稳定，训练集和测试集的RMSE分别为0.004 39和0.005 41，预测误差极低，说明随机森林模型对样本数据的预测精度较高，具备良好的泛化性能。

3.1.3　主控因素排序

基于优化后的随机森林模型，进一步利用特征重要性分析方法对影响CO_2-WAG驱开发效果的各主控因素进行了定量评估与深入分析，明确其相对重要性与作用程度。各主控因素重要性分析结果如图7所示。对于本研究的致密油藏算例，各主控因素的重要性排序为：注水时间百分比>注气速度>单轮次注入时间>注水速度>生产压差。表明在当前的地质和流体条件下，注水时间百分比是决定最终开发效果的首要因素，其次是注气速度。在致密储层中，合理的气水比例可以有效调控注入气体的运移路径，防止气体过早突破和窜流，改善驱替前缘稳定性。而维持较高的注入速度则有助于克服启动压力梯度，扩大波及体积。鉴于这一排序是开发参数与研究区块独特的地质物理特征（如高非均质性、低渗透基质、特定流体性质等）相互作用的综合结果，对于其他类型的储层，以上结论的适用性需进一步验证。

3.2　开发策略优化与最优生产预测

为明确致密油藏CO_2-WAG驱开发的最优生产方

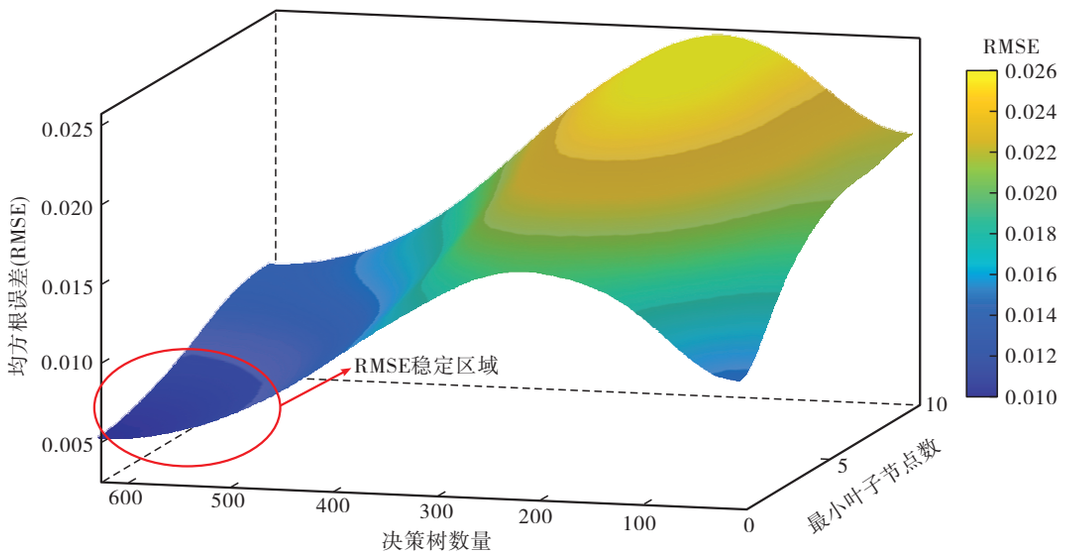


图4　超参数曲面拟合图

Fig. 4　Fitted surface diagram of hyperparameters

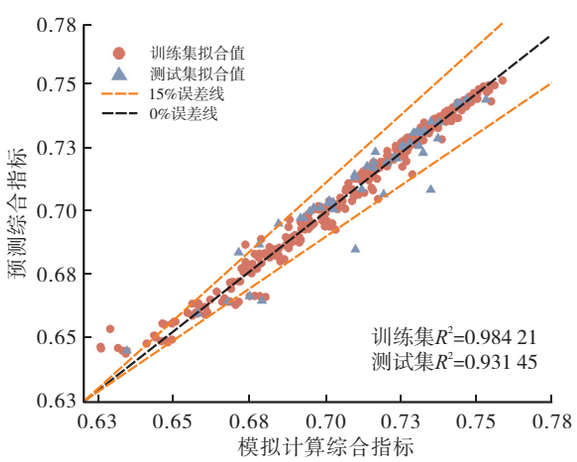


图5　优化后模型拟合效果图

Fig. 5　Fitting performance of the optimized model

案，基于优化后的随机森林模型，采用循环遍历法对各关键生产参数进行组合优化，生成大量参数组合方案并进行预测评价。通过对比分析最终确定了预测最优开发方案，见表3。经过参数组合优化后，预测全区最优方案在20年开发周期内平台累计增油量提高了8.6%，累计CO_2埋存量增加了14.5%。

以生产井WELL-P1为例进行模拟，分析样本最优方案与预测最优方案在实际开发效果上的差异，进一步验证优化方法的有效性，模拟结果如图8、图9所示。图8(a)中，在样本最优方案条件下，WELL-P1井周围区域的含油饱和度较高，表明局部驱替效果不理想，剩余油分布较为集中，波及范围有限，储层中大量剩余油未能得到有效动用。该现象说明，样本最优方案所采用的注采参数组合未能充分发挥储层潜力，导致驱替效率较低，进而影响油藏整体采收率。图8(b)中，在预测最优方案条件下，通过优化注水时间百分比和注气速度等关键生产参数显著改善

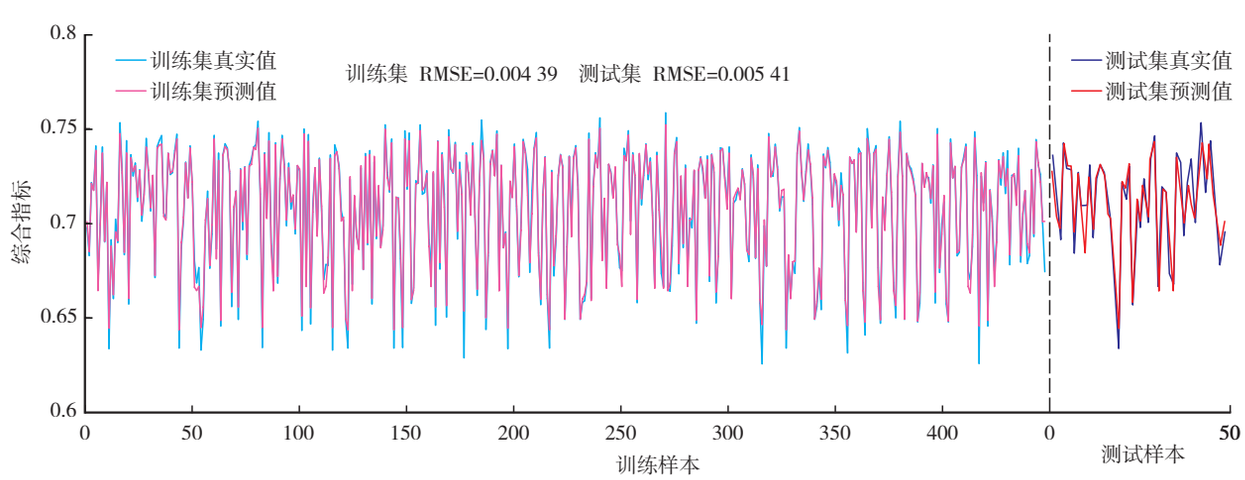


图6　优化后模型均方根误差图

Fig. 6　Root mean square error of the optimized model

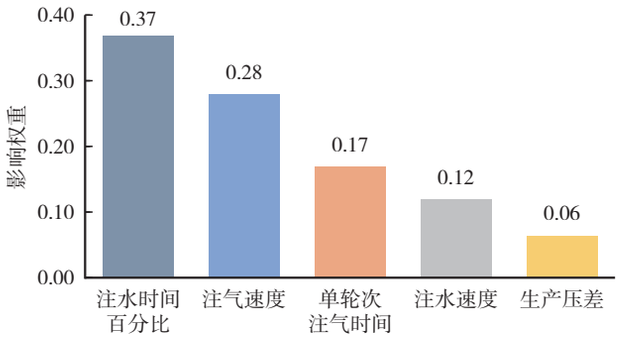


图7　优化后模型主控因素重要性排序图

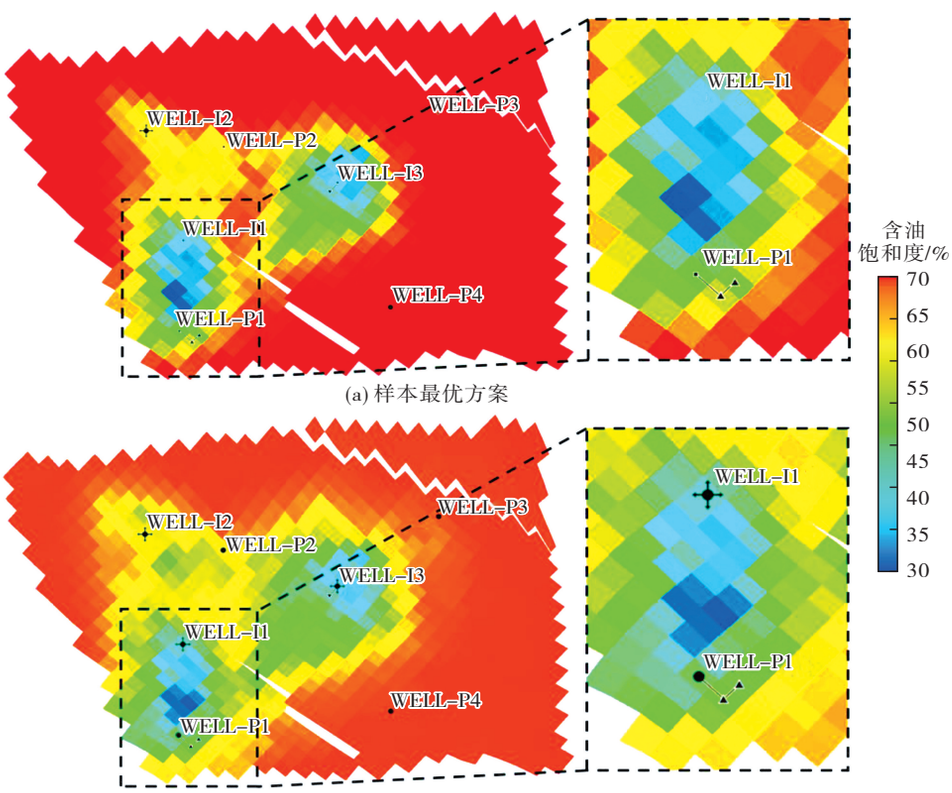
Fig. 7　Importance ranking of controlling factors based on the optimized model

了WELL-P1井区域的驱替效果。优化方案延长了注气时间段,同时调整注气速度和调整单轮次注入时间,以控制气体在储层中的流动路径,最终扩大了波及范围,提升了驱替均匀性,含油饱和度明显下降,剩余油得到有效动用。优化后的方案在更大程度上挖掘了油藏中剩余油的开发潜力,驱替波及范围更大,表明预测最优方案在储层动用程度和驱替效率方面具有显著优势。

表3　最优生产方案对比

Tab. 3　Comparison of optimal production schemes

参数	预测最优方案	样本最优方案
生产压差/MPa	14	18
注气速度/(10^3 m^3·d^{-1})	12	10
注水速度/(t·d^{-1})	35	40
注水时间百分比/%	14	25
单轮次注入时间/d	250	240
增油量/10^4 t	1.64	1.51
CO_2埋存量/10^6 m^3	29.2	25.5



(a) 样本最优方案

(b) 预测最优方案

图8　不同方案开发20年后含油饱和度对比

Fig. 8　Comparison of oil saturation after 20 years of development in cases of different schemes

从图9的模拟动态生产曲线可更直观地看出预测最优方案的效果。与样本最优方案相比,预测最优方案表现出更高的日产油量峰值,且高产稳产期显著延长。说明优化后的注采参数组合在短期内提升了油井产能,实现了更持久、更稳定的开发效果,降低了生产波动的风险。特别是在高产稳产期的延长方面,优化方案通过合理的注采控制有效延缓了储层压力的快速衰减,为长期开发提供了更稳定的保障。在气窜控制方面,预测最优方案同样展现出显著优势。通过对注水速度、注气速度以及注水时间百分比等参数的优化调整,有效控制了CO_2气体在储层中的流动路径和速度,延缓了气体在生产井处的突破时间。与样本最优方案相比,预测最优方案将CO_2气窜的发生时间推迟了近6年,充分体现了其在气体利用效率和驱替

控制能力方面的优越性。需要指出的是，该结果是基于数值模拟样本的对比分析，在实际生产过程中受实际地层条件、现场操作环境及生产配套设施的影响，实现如此大幅度的气窜延迟可能面临一定的实施难度，但优化方案在理论层面展现出显著的增油潜力和气体利用效率提升空间。

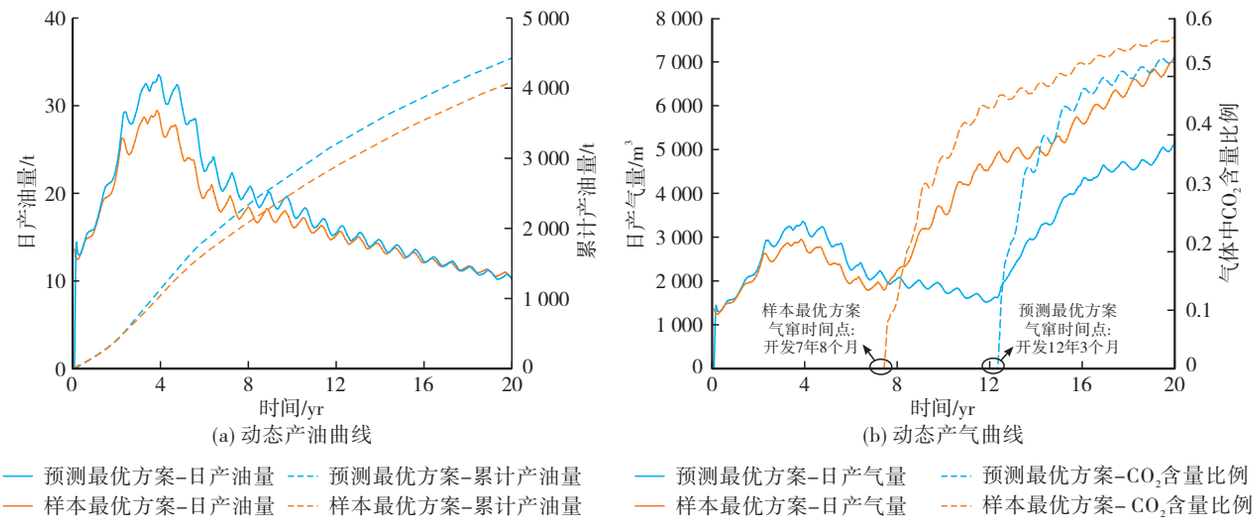


图9　WELL-P1井模拟动态生产曲线

Fig. 9　Simulated production curves of WELL-P1

为适应不同开发目标需求，进一步分析了权重系数（埋存率:增油量）对最优开发策略的影响，验证优化模型在不同目标导向下的灵活性与实用性。表4为不同权重系数下的优化方案对比，可通过调整权重系数以平衡CO_2埋存率与累计增油量的优先级，不同权重系数对优化结果的影响显著。当权重系数为0.75:0.25（偏向CO_2埋存率）时，优化方案的CO_2埋存率达到0.71，但累计增油量相对降低，为343 t；当权重系数为0.90:0.10（高度强调CO_2埋存率）时，优化方案的CO_2埋存率进一步提高至0.90，但累计增油量仅为126 t；当权重系数为0.10:0.90（偏向增油量）时，优化方案的累计增油量达到最大值，为762 t，但CO_2埋存率相对降低，为0.63。

表4　不同权重系数下优化方案对比

Tab. 4　Comparison of optimized schemes in cases of different weight factors

权重系数（埋存率:增油量）	埋存率	增油量/t
0.75:0.25	0.71	343
0.90:0.10	0.90	126
0.10:0.90	0.63	762

上述结果表明，权重系数的调整可以显著影响优化方案的开发效果，使其在CO_2埋存率和增油量之间实现灵活平衡。这种灵活性使得本文优化方法能够根据不同的开发目标（如增油优先或埋存优先）提出相应的最优开发策略，满足多样化的实际生产需求。

为了进一步验证优化模型预测结果的可靠性，通过数值模拟对预测最优方案进行了对比分析，结果如表5所示。以综合指标为评估参数（基于权重系数0.75:0.25），对比预测最优综合指标与数值模拟综合指标，结果显示预测值为0.767，数值模拟计算值为0.758，相对误差仅为0.9%。优化模型在较高的预测精度下能够有效捕捉输入参数与开发效果之间的复杂非线性关系。

表5　综合指标结果对比

Tab. 5　Comparison of comprehensive index results

参数	综合指标权重系数	预测最优综合指标	数值模拟综合指标	相对误差/%
数值	0.75:0.25	0.767	0.758	0.9

4　结论

1）通过改进的K-Fold交叉验证方法优化随机森林模型的超参数，确定了最优模型参数组合（最小叶子节点数为1、决策树数量为625、训练集与测试集比例为9:1）。优化后的模型表现出较高的拟合精度和稳定的泛化能力，测试集的预测误差控制在合理范围内。

2）基于随机森林模型的特征重要性分析方法，量化了关键参数对CO_2-WAG驱开发效果的影响程度。注水时间百分比是影响驱油效果的首要因素，其次是注气速度和单轮次注入时间。合理的注水时间百分比具有显著的增油效果，应作为开发方案设计的重点控制变量。

3）结合随机森林模型与循环遍历法，确定了致密

油藏CO_2-WAG驱的全局最优开发方案。与样本最优方案相比,预测优化方案可使累计增油量提高8.6%,CO_2埋存量增加14.5%,并有效推迟气体窜流时间近6年,显著改善了驱替效果和提高了开发效益。

4)基于随机森林模型的优化方法具备良好的灵活性,可根据不同的开发目标需求(如增油优先或埋存优先)调整权重系数,提出相应的最优开发策略。优化模型的预测结果与数值模拟结果高度一致,具有较高的预测精度和可靠性。

参考文献

[1]胡永乐,郝明强,陈国利,等. 中国CO_2驱油与埋存技术及实践[J]. 石油勘探与开发,2019,46(4):716-727.
HU Yongle, HAO Mingqiang, CHEN Guoli, et al. Technologies and practice of CO_2 flooding and sequestration in China [J]. Petroleum Exploration and Development, 2019, 46(4): 716-727.

[2]袁士义,马德胜,李军诗,等. 二氧化碳捕集、驱油与埋存产业化进展及前景展望[J]. 石油勘探与开发,2022,49(4):828-834.
YUAN Shiyi, MA Desheng, LI Junshi, et al. Progress and prospects of carbon dioxide capture, EOR-utilization and storage industrialization[J]. Petroleum Exploration and Development, 2022,49(4): 828-834.

[3]宋新民,王峰,马德胜,等. 中国石油二氧化碳捕集、驱油与埋存技术进展及展望[J]. 石油勘探与开发,2023,50(1):206-218.
SONG Xinmin, WANG Feng, MA Desheng, et al. Progress and prospect of carbon dioxide capture, utilization and storage in CNPC oilfields[J]. Petroleum Exploration and Development, 2023,50(1): 206-218.

[4]曹小朋,熊英,冯其红,等. 低渗透-致密油藏CO_2驱油与封存协同评价方法[J]. 油气地质与采收率,2023,30(2):44-52.
CAO Xiaopeng, XIONG Ying, FENG Qihong, et al. Collaborative evaluation method of CO_2 flooding and storage in low-permeability and tight reservoirs[J]. Petroleum Geology and Recovery Efficiency, 2023, 30(2): 44-52.

[5]武杨青,翟亮,鲁守飞,等. 碳中和背景下CO_2驱油技术研究进展[J]. 山东化工,2023,52(1):109-111.
WU Yangqing, ZHAI Liang, LU Shoufei, et al. Progress of CO_2 flooding technology in the context of carbon neutrality [J]. Shandong Chemical Industry, 2023, 52(1): 109-111.

[6]高慧梅,何应付,赵淑霞,等. 低渗透油藏CO_2驱气窜影响因素分析及模糊评判预测[J]. 中国海上油气,2014,26(4):63-66.
GAO Huimei, HE Yingfu, ZHAO Shuxia, et al. Control-factor analysis and fuzzy-evaluation prediction of gas channeling during CO_2 flooding in low-permeability oil reservoirs [J]. China Offshore Oil and Gas, 2014, 26(4): 63-66.

[7]李阳. 低渗透油藏CO_2驱提高采收率技术进展及展望[J]. 油气地质与采收率,2020,27(1):1-10.
LI Yang. Technical advancement and prospect for CO_2 flooding enhanced oil recovery in low permeability reservoirs [J]. Petroleum Geology and Recovery Efficiency, 2020, 27(1): 1-10.

[8]杜代军,方泽洲,章洋阳,等. 致密油藏CO_2驱响应性防气窜体系的制备及性能[J]. 精细化工,2024,41(5):1135-1142.
DU Daijun, FANG Zezhou, ZHANG Yangyang, et al. Preparation and properties of responsive channeling prevention system for tight oil reservoir CO_2 flooding[J]. Fine Chemicals, 2024, 41(5): 1135-1142.

[9]CHEN S, LI H, YANG D, et al. Optimal parametric design for water-alternating-gas (WAG) process in a CO_2-miscible flooding reservoir[J]. Journal of Canadian Petroleum Technology, 2010, 49(10): 75-82.

[10]FANG P, ZHANG Q, ZHOU C, et al. Chemical-assisted CO_2 water-alternating-gas injection for enhanced sweep efficiency in CO_2-EOR [J]. Molecules (Basel, Switzerland), 2024, 29(16): 3978.

[11]邹才能,杨智,董大忠,等. 非常规源岩层系油气形成分布与前景展望[J]. 地球科学,2022,47(5):1517-1533.
ZOU Caineng, YANG Zhi, DONG Dazhong, et al. Formation, distribution and prospect of unconventional hydrocarbons in source rock strata in China[J]. Earth Science, 2022, 47(5): 1517-1533.

[12]BAI Y, HOU J, LIU Y, et al. Energy-consumption calculation and optimization method of integrated system of injection-reservoir-production in high water-cut reservoir [J]. Energy, 2022, 239: 121961.

[13]AN Z, ZHOU K, HOU J, et al. Accelerating reservoir production optimization by combining reservoir engineering method with particle swarm optimization algorithm[J]. Journal of Petroleum Science and Engineering, 2022, 208: 109692.

[14]LI D, WANG X, XIE Y, et al. Analytical calculation method for development dynamics of water-flooding reservoir considering rock and fluid compressibility [J]. Geoenergy Science and Engineering, 2024, 242: 213250.

[15]HOURFAR F, SALAHSHOOR K, ZANBOURI H, et al. A systematic approach for modeling of waterflooding process in the presence of geological uncertainties in oil reservoirs [J]. Computers & Chemical Engineering, 2018, 111: 66-78.

[16]JABER A K, AL-JAWAD S N, ALHURAISHAWY A K. A review of proxy modeling applications in numerical reservoir simulation [J]. Arabian Journal of Geosciences, 2019, 12 (22): 701.

[17]MAO S, MEHANA M, HUANG T, et al. Strategies for hydrogen storage in a depleted sandstone reservoir from the San Joaquin Basin, California (USA) based on high-fidelity numerical simulations[J]. Journal of Energy Storage, 2024, 94: 112508.

[18]蒲万芬,靳星,唐晓东,等. 基于BP神经网络的低渗透底水油藏油井见水模式预测模型[J]. 新疆石油天然气,2024,20(2):37-47.
PU Wanfen, JIN Xin, TANG Xiaodong, et al. Prediction model of water breakthrough patterns of low-permeability bottom water reservoirs based on BP neural network[J]. Xinjiang Oil & Gas,2024,20(2): 37-47.

[19]MAHDAVIARA M,SHARIFI M,AHMADI M. Toward evaluation and screening of the enhanced oil recovery scenarios for low permeability reservoirs using statistical and machine learning techniques[J]. Fuel,2022,325: 124795.

[20]MENG S,FU Q,TAO J,et al. Predicting CO_2-EOR and storage in low-permeability reservoirs with deep learning-based surrogate flow models[J]. Geoenergy Science and Engineering,2024,233: 212467.

[21]GONG A,ZHANG L. Deep learning-based approach for reservoir fluid identification in low-porosity, low-permeability reservoirs[J]. Physics of Fluids,2025,37 (4):046611.

[22]李宏波,罗平亚,白杨,等. 机器学习算法概述及其在钻井工程中的应用[J]. 新疆石油天然气,2022,18(1):1-13.
LI Hongbo, LUO Pingya, BAI Yang, et al. Summary for machine learning algorithms and their applications in drilling engineering[J]. Xinjiang Oil & Gas,2022,18(1):1-13.

[23]ZHOU W,LIU C,LIU Y,et al. Machine learning in reservoir engineering: a review[J]. Processes,2024,12 (6): 1219.

[24]YOU J, AMPOMAH W, SUN Q, et al. Machine learning based co-optimization of carbon dioxide sequestration and oil recovery in CO_2-EOR project[J]. Journal of Cleaner Production,2020,26: 120866.

[25]VO THANH H, SHEINI DASHTGOLI D, ZHANG H, et al. Machine-learning-based prediction of oil recovery factor for experimental CO_2-foam chemical EOR: Implications for carbon utilization projects[J]. Energy,2023,278: 127860.

[26]LI H,GONG C,LIU S,et al. Machine learning-assisted prediction of oil production and CO_2 storage effect in CO_2-water-alternating-gas injection(CO_2-WAG)[J]. Applied Sciences,2022,12 (21): 10958.

[27]KANAANI M, SEDAGHAT A M, AMARZA DEH A, et al. Stacking learning for smart proxy modeling in CO_2-WAG optimization: A techno-economic approach to sustainable enhanced oil recovery[J]. ACS Omega,2025,10 (9),9563-9582.

[28]熊玥雯,向海昀,杨宏楠. 随机森林算法在石油领域的应用[C]//西安石油大学,陕西省石油学会. 2024油气田勘探与开发国际会议论文集Ⅰ,陕西西安:西南石油大学计算机与软件学院,西南石油大学石油与天然气工程学院,2024:616-626.
XIONG Yuewen,XIANG Haiyun,YANG Hongnan. Application of random forest algorithm in petroleum [C]//Xi'an Shiyou University,Shaanxi Petroleum Society. Proceedings of the 2024 International Conference on Oil and Gas Field Exploration and Development I,Xi'an,Shaanxi,China: School of Computer and Software,Southwest Petroleum University,School of Petroleum and Natural Gas Engineering,Southwest Petroleum University,2024: 616-626.

[29]方匡南,吴见彬,朱建平,等. 随机森林方法研究综述[J]. 统计与信息论坛,2011,26(3): 32-38.
FANG Kuangnan,WU Jianbin,ZHU Jianping,et al. A review of technologies on random forests [J]. Journal of Statistics and Information,2011,26(3): 32-38.

[30]李正良,彭思思,王涛. 基于k-fold交叉验证的代理模型序列采样方法[J]. 计算力学学报,2022,39(2):244-249.
LI Zhengliang, PENG Sisi, WANG Tao. A sequential sampling method of surrogate model based on k-fold cross validation[J]. Chinese Journal of Computational Mechanics,2022,39(2): 244-249.

[31]LOPEZ-ECHEVERRY J S,REIF-ACHERMAN S,ARAUJO-LOPEZ E. Peng-Robinson equation of state: 40 years through cubics[J]. Fluid Phase Equilibria,2017,447: 39-71.

[32]CENTENO G,SÁNCHEZ-REYNA G,ANCHEYTA J,et al. Testing various mixing rules for calculation of viscosity of petroleum blends[J]. Fuel,2011,90(12): 3561-3570.

[33]MORTAZAVI-MANESH S,SATYRO M,MARRIOTT R A. A semi-empirical Henry's law expression for carbon dioxide dissolution in ionic liquids[J]. Fluid Phase Equilibria,2011,307(2): 208-215.

[34]BIAU G. Analysis of a random forests model[J]. The Journal of Machine Learning Research,2012,13: 1063-1095.

(修改回稿日期:2025-12-20　编辑:何园)